DOCUMENT RESUME

ED 064 404 TM 001 652

AUTHOR Willard, Richard W.
TITLE The Nationwide Survey.
PUB DATE 72
NOTE 10p.; Paper presented at annual meeting of the AERA for a symposium on "Some Problems Associated with Nationwide Evaluation of Educational Problems in Schools," (Chicago, Ill., April 1972)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Data Collection; Decision Making; *Federal Programs; *National Surveys; Objectives; Organization; *Policy Formation; Role Perception; School Districts; *Surveys

ABSTRACT

Evaluation in education has come to be seen as an essential ingredient in educational decision making. Decisions to be made by educators cover a wide spectrum, varying according to the role of the teacher, principal, and superintendent. Education is essentially discussed at the Federal level and the nature of surveys, which is the sole efficient way of assessing the characteristics and needs of the countless school districts in the United States is discussed. The design or structure of a survey depends first of all on its intended objectives, i.e., the types of questions it hopes its respondents will answer. A type of pyramiding is described for conducting federal-state programs. The use of the survey to collect other than routine data is discussed. It is concluded that because of the limitations inherent with surveys, they should be used mainly as a means to answer a few simple policy questions that require data that can be collected reasonably accurately without creating undue response burdens. (CK)

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

21.16
10 p

ED 064404

THE NATIONWIDE SURVEY

Richard W. Willard
New England School Development Council

TM 001 652

Prepared for
a symposium on "Some Problems Associated with Nationwide Evaluation of Educational Problems in Schools," 1972 Annual Meeting of the American Educational Research Association

While there has long been a concern for evaluation of educational programs, only in the past five years or so has there evolved a definition of evaluation that is generally accepted. In prior years there has been a confusion between research and evaluation due primarily to the fact that both the researcher and the evaluator use statistics in their work. The difference that was missed lies, of course, in the intended use of statistical inference. For instance the researcher's intention is to establish or reaffirm a truth, while the evaluator is concerned with supporting or enhancing some decision-making process. Thus, evaluation in education--whatever model one chooses to use--has come to be seen as an essential ingredient in educational decision-making.

The decisions to be made by educators cover a wide spectrum, being as varied as the roles people play in the educational enterprise, and therefore evaluation serves a crucial function at all levels. The teacher, for example, will appreciate evaluative feedback on student progress in the middle of an individually prescribed program or an independent study program. The principal of an open classroom school will welcome evaluative information regarding the use of resource material or the differential effects on his students of various degrees of classroom structure. The superintendent from his level of responsibility will depend upon evaluation almost daily as he makes or recommends decisions concerning budgets, continuation of programs, hiring and firing of personnel, the assessment of student needs, or the setting and resetting of program objectives.

We can move away from the local environment to consider needs for evaluation at the state level. A state officer may well--and probably does--hope for evaluation that would identify, for example, which models for Title I reading projects work best in his state. The implications are obvious in that he can then be in a position to recommend to his LEA's several proven approaches.

In this progression we come eventually to the question : what kind of evaluation needs has the Federal government, particularly the U.S. Office of Education? What are the kinds of decisions that the USOE must make about educational programs? We observe, first, that these decisions, which have primarily to do with the programs that enjoy full or partial Federal support, relate to two quite different concerns. First, the USOE has the responsibility of reporting to the Congress about the present status of programs and of recommending changes in educational legislation. Such recommendations take many forms, such as changes in formulas used to determine the delivery of dollars or even the cessation of a particular Federal program, or, there can be changes in the legislation that shift the emphases between or among programs. A second concern relates to USOE's responsibility for making recommendations or creating guidelines that help states and locals to formulate their projects. In this context the USOE has some of the same evaluative concerns as do the states, namely, what models for projects seem to work best under what conditions? Knowing such information and passing it along appropriately should help those at state and local levels to create more successful projects.

Given such concerns, what kinds of information can the USOE most reasonably collect? Well, the legislation now calls for the use of objective measurements by the local schools in evaluating their programs. It was left to the local schools to develop their own evaluation studies, and this seems very reasonable. After all, the local schools are supposed to conduct an assessment to determine the special educational needs of their children. Following the assessment, local schools are responsible for setting their own program objectives and for employing the appropriate measures for program evaluation.

If evaluations should be designed and executed by the local schools, why haven't the states and the USOE been able to collate the results of local evaluation in order to create a nationwide summary? It seems clear that such an effort has not worked mostly because local reports are so subjective, but we can only conjecture about the probable causes for that. Among other problems, though, I must note the problem of combining in any meaningful way the many different kinds of local programs. Because reading

tends to be a common problem attacked most often by compensatory education programs it provides a good example of this problem. Not only can reading programs differ from grade to grade, but their contents can differ in more ways than they are typically described. In some schools the Title I label might be affixed to all reading activities. In others, only the remediation work is included under Title I. Without careful scrutiny we may never know the extent to which a reading program is individualized or uses tutors and of what kind. There are just a few of the kinds of attributes of reading programs that make it difficult to know how to group local programs in any meaningful way in order to summarize the results of local evaluations.

Another problem involving summary analyses of locally determined objective measures arises with those problems that have objectives taken from the affective domain. While there are many measures to choose from for dealing with the cognitive domain, there are not really many appropriate for the affective domain. And many of those require the techniques of systematic observation, which in turn require special training. This suggests a third point, namely, that local educational agencies do not all have the expertise to conduct proper program evaluation. It is not a small wonder, then, that many early local evaluation reports contained very subjective testimonials to the success of their programs and left it at that.

Whatever the reasons, however, it seems to most observers that summarizing local evaluation reports is inadequate for purposes of nationwide evaluation. We must look for alternative approaches. My responsibility today is to look at the nationwide survey as one of these alternatives.

In looking at the nationwide survey I will be interested in two major attributes of surveys. The first of these concerns structure. The survey uses a structured instrument of instructions and questions that guide the collection of data. The survey also uses a structured set of respondents in order to have a sample that provides some desired kind of representation. In addition, a good survey design should include a structured plan for the analysis of data.

A survey should have, most of all, a structure that relates the general objectives of the survey to the data being collected. Now, there is a group

of us at NESDEC that has had some experience with such a structure while we were serving as a contractor to the USOE to develop the instrumentation for a nationwide survey of secondary schools for which the goal is evaluative information about Federally supported programs in secondary schools. Working with the Joint Federal/State Task Force on Evaluation we have developed a structure approach that I want to share with you.

One common first step in the design of a survey is to put right into the instruments the questions it is hoped the survey will answer. For example, poll surveys ask the individuals to record their voting preferences. This approach is not totally appropriate for the kind of survey we are dealing with because, for example, one cannot ask the simple direct question: "Did the funds reach the targetted population?" Some respondents won't even know what that question means. When the respondent does know, it suggests simply a yes/no answer without any indication of extent. In fact, the better question for policy considerations is: "To what extent are Title I funds appropriately targetted?" Even that question doesn't suggest directly what the instruments should collect. Another, lower level, question or two is needed. Examples of those are as follows:

I. To what extent are school districts with the highest concentration of pupils from low-income families and the greatest relative need receiving an equitable share of Title I funds?

This question deals with the selection of schools districts for Title I aid as well as with the relationship between the degree of need and level of aid. We see that to answer this question we need to know for any given district the number of children in the attendance area and, among those, the number from low-income families. In fact, we can imagine another, lower level of questions that indicate the data requirements. One such question, then, requires the amount of Title I funding by district. You can see, thus far, that lower level questions tend to grow in number in order to answer a higher level question. We can also see this as we return to another question that will help answer our main question about the targetting of funds. Thus far we have dealt with districts. Now, let us consider concerns for the selection of schools with the following question:

II. To what extent are schools with the highest concentration of pupils from low-income families designated as Title I schools?

This one question suggests the requirement for having data about presence or absence of Title I as well as about the proportion of children from low-income families, school by school. Consider, now, a third major question concerning targetting.

III. To what extent are the most educationally deprived pupils selected to participate in Title I services?

This question clearly deals with the decisions that involve individual pupils. It can be considered as calling for the comparison of selected participants with non-participants, and that comparison can be further clarified by more lower level questions. In another vein, the same question can also be construed to deal with the types of selection procedures used, and that can be covered with another, lower level question.

I hope that my description of a hierarchy of questions has been sufficiently clear for you to see the pattern evolving. For a given policy concern it is possible to affix lower levels of policy concerns and, eventually, data requirements in a pyramid appearing structure. The question at the very top level of the pyramid is about a very general concern. Moving down the pyramid, questions become more specific until at the very base of the pyramid they express data requirements. The lowest level questions either can appear as is on questionnaires, when that is appropriate, or they can suggest variables that should be derived from questionnaire items. Even though I have outlined only a part of the process, I think that you should begin to see, now, one important component of the approach we have used to create pyramids of questions. The pyramids provide a linkage between instrument variables and the policy questions or policy areas.

Our group at NESDEC prepared many, many sets of pyramids in behalf of policy concerns that had been expressed--albeit in less detail--previously by state and Federal officials. After we had prepared the pyramids, their importance suggested that there should be more than a simple review and revision of them. Accordingly, the Joint Federal/State Task Force on Evaluation worked out a complex review procedure that provided an opportunity for minority

positions to be heard rather than smothered. It came to be called a modified Delphi approach, and I'll describe only briefly how it worked. Materials were distributed to a fairly large number of state and Federal officers, including among the latter, some who serve on Congressional staffs. Each person was directed to review, rate, and rank the policy questions down to the third level of the pyramids. A summary of the ratings and rankings went back to each person along with his original reply in order that he could see his own ranking in contrast to all others. He could then, you see, decide to alter his position to fit the rest or conduct a concerted effort to convince the others of his position. Such moves were made in committee meetings devoted expressly to resolving differences and to deriving one final, compromise set of ratings and rankings.

A review of lower level questions--those that suggest the data requirements--was handled separately in order to determine whether a question should be included or excluded and whether new data questions should be added.

Let me describe, briefly, the magnitudes of the pyramids we have been creating. First of all, we have done the task for more than just ESEA, Title I. We have worked, also, on ESEA, Titles II, III, VI, VII, VIII, on the Vocational Education Amendments of 1968 and on NDEA, Title III. The number of pyramids varies a little from Title to Title. For Title I we ended up with four pyramids after the review process. The key questions for each of the four pyramids are as follows:

A. To what extent are Title I funds appropriately targetted?
B. Are services addressed to the special educational needs of the participants?
C. What effects are associated with Title I services?
D. Is there a need for change in the Federal and state conduct of Title I?

To give some idea of the shape of the pyramids, at the next level down from the four questions above for Title I there were a total of ten questions. And, the next level down from that had thirty-four questions. It becomes difficult to count beyond that level due to the fact that the questions begin to get closer and closer to specifying variables and any one variable might

appear under several different hierarchies according to its relevance. Such redundancy, in fact, serves to indicate the overall importance of a given variable. When it comes time, inevitably, to place priorities on an unduly large set of questionnaire items, those repeated least often are the best candidates for deletion. Further, knowing which policy concern they address allows for the deletion process being a rationale process.

The fact that variable specifications are imbedded in a hierarchy of questions proved to us to make very straightforward the general plan for data analysis. Coupled together were the needed variables as well as some indication of the desired analysis, whether it be generally a univariate frequency distribution or a cross-tabulation of either a simple or complex nature. This process worked well with the first two of the four pyramids, those dealing with targetting and with appropriateness of services. The fourth area was similarly amenable to this approach. The third area--that of the effects of various services--proved to be a problem, however.

While looking at the potential data analysis plans that could be relevant for answering questions about effect that fell under that pyramid, we came to the conclusion that the survey approach is inappropriate. We came to question the extent to which a single survey could approach even the weakest of the various post hoc experimental designs. Far better would be some form of classical experimental design.

Thus the use of pyramids, putting in perspective as they do the demands for data as well as the data analysis requirements, made it more logical to defer questions dealing with effects to a different and more appropriate approach. Our use of pyramids, then, not only provided a structure on which to build the overall survey effort, but as well it allowed us to come to more realistic terms with the limitation of the survey.

I turn now to the second attribute of surveys with which I will deal, namely that a survey is an event that is outside the ordinary range of business. As such, the survey can be used to collect data that supplement other data that are routinely collected by forms that are, in fact, part of the regular business. Consider as an example a local Title I project. Certain reports to the state are routinely completed that provide data about participants.

We need an out of the ordinary event such as a survey, though, to supplement such data with other data about non-participants, say. Or, a survey can be used to collect some factual data otherwise not collected about Federally funded services in schools.

But, what is the price we pay for such benefits? Simply because it is out of the ordinary, a survey places a burden on its respondents. The longer and more complex a questionnaire is, the worse the response burden becomes. An unfortunate consequence of burdens on respondents that affect their cooperation is inaccuracies in the data collected. It seems that unreliability is the survey's constant companion. Christopher Jencks in Mosteller and Moynihan (1972)* has done a masterful job of sleuthing through some of the Coleman Survey data to derive estimates of the reliability of the data by checking out the consistency of responses. It is quite disconcerting to find that a lack of reliability crept into responses by principals about which is the lowest grade in their schools. The responses were unusually inconsistent with other responses about the presence and costs of kindergartens and nursery schools. In another case, Jencks found inconsistencies about the reporting of the presence or absence of a room set aside as a centralized school library.

Now, I realize that survey designers recognize some of the attributes of a survey that are likely to lead to inaccurate data. Generally the designer takes care to avoid the problems as much as he can. He tries to minimize the burden imposed upon the respondent by keeping the instrumentation as short and as simple as he can. He tries not to use any techniques that might alienate the respondent. For example, a survey designer might elect not to ask a parent the income of the family but might make a compromise and rely upon a pupil respondent to guess his family's income or maybe even ask someone in the school to make such a guess. This use of alternate respondents, you can readily see, results in a self-imposed inaccuracy in order to avoid another kind of inaccuracy. Neither kind of inaccuracy is tolerable, however.

Because there are such limitations inherent with surveys, my attitude is that we must learn to constrain ourselves from indiscriminate use of the

---

*Mosteller, Frederick and Moynihan, Daniel P., eds., On Equality of Educational Opportunity. Random House: New York, 1972.

survey approach. We should turn to the survey as a means to answer a few simple policy questions that require data that can be collected reasonably accurately without creating undue response burdens. By limiting ourselves to simple questions and simple data that we can reasonably deal with, we can avoid many of the negative aspects of surveys.

For those policy issues that suggest the level of complexity of, say, the Coleman Study there are reasons to look for alternate approaches. Using observation techniques, for example, would ensure better accuracy in reports of centralized school libraries or of lowest grades in a school. This approach perforce would lead to smaller samples than possible by survey, but the increase in accuracy and reduction in compromises seem to me to make that trade-off worthwhile.